Lesseur, C., Diergaarde, B., Olshan, A. F., Wünsch-Filho, V., Ness, A. R., Liu, G., Lacko, M., Eluf-Neto, J., Franceschi, S., Lagiou, P., Macfarlane, G. J., Richiardi, L., Boccia, S., Polesel, J., Kjaerheim, K., Zaridze, D., Johansson, M., Menezes, A. M., Curado, M. P., ... Brennan, P. (2016). Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nature Genetics*, *48*(12), 1544–1550. https://doi.org/10.1038/ng.3685

Peer reviewed version

Link to published version (if available):
10.1038/ng.3685

Genome-wide association study of oral cavity and pharyngeal cancer

Corina Lesseur[1], Brenda Diergaarde[2], Andrew F Olshan[3], Victor Wünsch-Filho[4], Andrew R Ness[5], Geoffrey Liu[6], Martin Lacko[7], José Eluf-Neto[8], Silvia Franceschi[1], Pagona Lagiou[10], Gary J Macfarlane[11], Lorenzo Richiardi[12], Stefania Boccia[13], Jerry Polesel[14], Kristina Kjaerheim[15], David Zaridze[16], Mattias Johansson[1], Ana M Menezes[17], Maria Paula Curado[18], Max Robinson[19], Wolfgang Ahrens[20], Cristina Canova[21], Ariana Znaor[1,22], Xavier Castellsagué[§23], David I Conway[24], Ivana Holcátová[25], Dana Mates[26], Marta Vilensky[27], Claire Healy[28], Neonila Szeszenia-Dąbrowska[29], Eleonóra Fabiánová[30], Jolanta Lissowska[31], Jennifer R Grandis[32], Mark C Weissler[33], Eloiza H Tajara[34], Fabio D Nunes[35], Marcos B de Carvalho[36], Steve Thomas[37], Rayjean J Hung[38], Wilbert H M Peters[39], Rolando Herrero[1], Gabriella Cadoni[40], H Bas Bueno-de-Mesquita[41,42,43], Steffen Annika[44], Antonio Agudo[45], Oxana Shan'gina[16], Xiangjun Xiao[46], Valérie Gaborieau[1], Amélie Chabrier[1], Devasena Anantharaman[1], Paolo Boffetta[47], Christopher I Amos[46], James D McKay[1], Paul Brennan*[1]

Affiliations

[1]International Agency for Research on Cancer (IARC/WHO), Lyon, France.

[2]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA.

[3]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, and UNC Lineberger Comprehensive Cancer Center, Chapel Hill, NC USA.

[4]Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, SP, Brazil.

[5]National Institute for Health Research (NIHR) Biomedical Research Unit in Nutrition Diet and Lifestyle at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol and School of Oral and Dental Sciences, University of Bristol, Bristol, UK.

[6]Princess Margaret Cancer Centre, Toronto, Canada.

[7]Department of Otorhinolaryngology, Head and Neck Surgery, Maastricht University Medical Center, Maastricht, Netherlands.

[8]Departamento de Medicina Preventiva, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

[10]Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Greece.

[11]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK.

[12]Department of Medical Sciences, University of Turin, Turin, Italy.

[13]Section of Hygiene, Institute of Public Health, Università Cattolica del Sacro Cuore, Fondazione Policlinico ‘Agostino Gemelli’, Rome, Italy.

[14]Unit of Cancer Epidemiology, CRO Aviano National Cancer Institute, Aviano, Italy

[15]Cancer Registry of Norway, Oslo, Norway.
[16]Department of Cancer Epidemiology and Prevention, Institute of Carcinogenesis, N.N. Blokhin Russian Cancer Research Centre of the Russian Ministry of Health, Moscow, Russian Federation.
[17]Universidade Federal de Pelotas (UFPel), Pelotas, RS, Brazil
[18]Epidemiology, International Center for Research (CIPE), A.C.Camargo Cancer Center, Sao Paulo, Brazil
[19]Centre for Oral Health Research, Newcastle University, Newcastle, UK.
[20]Leibniz Institute for Prevention Research and Epidemiology (BIPS), Bremen, Germany.
[21]Deparment of Molecular Medicine, University of Padova, Padova, Italy.
[22]Croatian National Institute of Public Health, Zagreb, Croatia.
[23]Institut Català d'Oncologia (ICO)-DIBELL, CIBER-ESP, L'Hospitalet de Llobregat, Catalonia, Spain.
[24]University of Glasgow, Glasgow, and NHS NSS, Edinburgh, UK.
[25]Institute of Hygiene & Epidemiology 1st Faculty of Medicine, Charles University, Prague, Czech Republic.
[26]National Institute of Public Health, Bucharest, Romania.
[27]Instituto de Oncologia "Angel H Roffo", Universidad de Buenos Aires, Buenos Aires, Argentina.
[28]Trinity College School of Dental Science, Dublin, Ireland.
[29]Department of Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland.
[30]Regional Authority of Public Health, Banská Bystrica, Slovakia.
[31]The Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology (MCMCC), Warsaw, Poland.
[32]Department of Otolaryngology Head and Neck Surgery, and Clinical and Translational Science Institute, University of California at San Francisco, San Francisco, CA, USA.
[33]Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
[34]Department of Molecular Biology, School of Medicine of São José do Rio Preto, São José do Rio Preto, SP, Brazil.
[35]Department of Stomatology, School of Dentistry, University of São Paulo, São Paulo, SP, Brazil.
[36]Department of Head and Neck Surgery, Heliópolis Hospital, São Paulo, SP, Brazil
[37]School of Oral and Dental Sciences, University of Bristol, Bristol, UK.
[38]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada.
[39]Department of Gastroenterology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands.
[40]Institute of Otorhinolaryngology, Università Cattolica del Sacro Cuore, Fondazione Policlinico 'Agostino Gemelli', Rome, Italy.
[41]Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands.
[42]Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, London, UK.

[43]Department of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.
[44]German Institute of Human Nutrition in Potsdam-Rehbruecke (DIfE), Nuthetal, Germany.
[45]Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain.
[46]Department of Community and Family Medicine Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA.
[47]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[§] Xavier Castellsagué recently passed away (6/12/2016)

***Corresponding Author:**

Paul Brennan, Ph.D.
Section of Genetics
International Agency for Research on Cancer (IARC)
Lyon, France
gep@iarc.fr

We conducted a genome-wide association study of oral cavity and pharyngeal cancer in 6,034 cases and 6,585 controls from Europe, North America and South America. We evaluated >7 million SNPs with oral and pharyngeal cancer risk and detected 8 loci at $P<5x10^{-8}$, 7 of which are novel for these cancer sites. Overall, oral and pharyngeal cancers were associated with novel loci at 6p21.32 (rs3828805, *HLA-DQB1*), 10q26.13 (rs201982221, *LHHP*) and 11p15.4 (rs1453414, *OR52N2/TRIM5).* Oral cancer was associated with two new regions 2p23.3 (rs6547741, *GPN1*) and 9q34.12 (rs928674, *LAMC3*), and with known cancer susceptibility loci 9p21.3 (rs8181047, *CDKN2B-AS1*) and 5p15.33 (rs10462706, *CLPTM1L*). Novel oropharyngeal cancer associations were limited to the human leukocyte antigen (HLA) region at 6p21.32 and HLA classical allele imputation revealed a protective association with the class II haplotype *DRB1*1301-DQA1*0103-DQB1*0603* (odds ratio (OR)=0.59, $P=2.7x10^{-9}$). Stratified analyses on a subgroup of oropharyngeal cases with human papillomavirus (HPV) infection status indicated that this association was considerably stronger in HPV-positive (OR=0.23, $P=1.6x10^{-6}$) compared to HPV-negative cancers (OR=0.75, $P=0.16$).

Cancers of the oral cavity (OC) and oropharynx (OPC) are predominantly caused by tobacco and alcohol use, although oral infection with HPV, particularly HPV16, is an increasingly important cause of OPC[1], especially in the US and northern Europe[1,2]. The proportion of HPV-related OPCs varies widely and is estimated to be approximately 60% in the US, 30% in Europe and lower in South America[2-5]. Genetic factors have also been implicated in OC and OPC susceptibility, especially polymorphisms within alcohol-related genes including alcohol-dehydrogenase 1B (*ADH1B*) and *ADH7*[6,7]. In order to identify additional susceptibility loci, 13,107 individuals from 12 epidemiological studies (**Supplementary Table 1**) were genotyped using the Illumina OncoArray and after stringent quality-control steps (**Supplementary Table 2, Online Methods**) 6,034 cases and 6,585 cancer-free controls remained for analyses (**Table 1**). We next performed genome-wide imputation using the Haplotype Reference Consortium panel[8] and obtained approximately 7 million high-quality imputed variants (**Supplementary Fig. 1)**. Given the ethnic diversity of our study, we evaluated associations within continent (Europe, North and South America) using multivariate unconditional logistic regressions under a log-additive genetic model adjusted for age, sex and regional eigenvectors. Results by continent were combined using fixed-effect meta-analyses to derive associations for overall OC and pharynx cancer (oral, oropharynx, hypopharynx and overlapping cancers; n=6,034), as well as site-specific OC (n=2,990) and OPC (n=2,641). Although, several ethnic groups are present in the study, supervised ancestry analyses indicated that >90% of participants were predominantly of Caucasian (>70%CEU) ancestry, although some population admixture was observed in South America (**Supplementary Table 3**).

GWA meta-analyses of overall and site-specific cancers identified 9 regions at genome-wide significance ($P<5x10^{-8}$) (**Fig.1**). Quantile-quantile (Q-Q) plots of observed and expected *P*-values showed moderate genomic inflation (λ) for the 3 meta-analyses (λ range=1.04-1.06, **Supplementary Fig. 2-3**). Since λ increases with sample size, we scaled it to 1000 cases and controls resulting in ameliorated inflation (λ_{1000} range=1.009-1.01)[9]. Overall OC and pharynx cancer were associated with rs79767424 (5p14.3), rs1229984 (4q23), rs201982221 (10q26.13), rs1453414 (11p15.4) and 123 SNPs at 6p21.32 (**Supplementary Table 4**). Twenty-six variants were associated ($P<5x10^{-8}$) with OC (**Supplementary Table 5**), 4 of which mapped to 2p23.3, 1 to 4q23, 3 to 9q34.12, 13 to 5p15.33 and 5 to 9p21.3. For OPC, novel significant variants were located at 6p21.32 (62 SNPs, **Supplementary Table 6**). Suggestive susceptibility variants ($P<5x10^{-7}$, **Supplementary Tables 7-9**) were associated with OC at 4 additional loci: 6p21.33, 6p21.32, 15q21.2, 15q26.2 and, with OPC at 2q36.1. Other genomic locations outside the HLA region showed promising associations ($P<5x10^{-6}$) with OPC (**Supplementary Table 10**). For susceptibility loci at $P<5x10^{-8}$, functional annotation of regulatory features with ENCODE and eQTL information, if available, are summarized in **Supplementary Table 11 and 12.** Given the geographical heterogeneity of our population, we performed sensitivity analyses after excluding individuals with <70% CEU ancestry and these showed similar results (**Supplementary Table 13**). To validate array genotypes and imputed dosages, we directly genotyped by a different platform (TaqMan) at least one variant within each locus ($P=5x10^{-7}$) in a subset of approximately 700 individuals. Concordance between genotyped/imputed genotypes and TaqMan results was >97% for all regions with the exception of

rs2398180, an imputed variant which had a concordance of 94% (**Supplementary Table 14**). For 2 rare variants, rs201982221 (10q26.13) and rs7976742 (5p14.3), TaqMan assays could not be designed and we used Sanger sequencing for validation. We were able to validate the rs201982221 deletion (**Supplementary Fig. 4)**, but rs7976742 did not validate (**Online Methods**). The lead variant at each validated loci ($P<5x10^{-8}$) for overall and site-specific analyses are shown in **Table 2**.

The rs1229984 (4q23, *ADH1B*) association has been previously reported as a susceptibility locus for OC and OPC[6], and similar to previous findings this variant showed heterogeneity by region, smoking and alcohol drinking status[10,11] (**Fig. 2a**). Three other 4q23 SNPs reached $P<5x10^{-8}$, although conditional analyses indicated these are not independent signals (**Supplementary Table 15**). The rs1573496 (*ADH7*) variant reported to be strongly associated with OC and OPC in the previous upper aerodigestive tract cancer GWAS[7] was only moderately associated here (**Supplementary Table 16**). In the overall OC and pharynx cancer analysis, we identified rs201982221 at 10q26.13 (OR=1.67, $P=1.58x10^{-9}$), that was also separately associated with OC (OR=1.71, $P=1.04x10^{-7}$) and OPC (OR=1.70, $P=7.9x10^{-7}$) (**Fig. 2b**). rs201982221 is located within the *LHPP* gene in a region with reported regulatory features (**Supplementary Table 11**). However, it is a rare intronic deletion in an area of low linkage disequilibrium (LD) (**Supplementary Fig. 5**), and thus warrants further validation in a different population. rs1453414, the lead signal at 11p15.4 (**Supplementary Table 17**), is an intronic variant that showed a borderline association in the overall (OR=1.19, $P=4.78x10^{-8}$) and site-specific analyses [OC (OR=1.19, $P=1.65x10^{-5}$) and OPC (OR=1.22, $P=4.26x10^{-6}$)] (**Fig. 2c, Supplemental Fig. 6**).

rs1453414 is upstream of *OR52N2,* an olfactory receptor, and within *TRIM5,* an E3-ubiquitin ligase, and is an eQTL for these genes in brain tissue[12] (**Supplementary Table 12)**.

At 2p23.3, 4 SNPs showed evidence ($P<5x10^{-8}$) for an association with OC, and in conditional analyses did not appear to be independent (**Supplementary Table 18**). These signals map to a high LD area that includes *C2orf16, ZNF512, CCDC121* and *GPN1* (**Supplementary Fig. 7**). The lead SNP, rs6547741, was associated with OC but not with OPC, and maps to an intron of *GPN1*, a GTPase involved in RNA polymerase II transport and DNA repair[13]. Associations between rs6547741 and OC were homogenous across other stratified analyses by region, sex, smoking and drinking status (**Fig. 3a**).

Variation within 5p15.33 was also exclusively associated with OC (OPC, rs10462706, P=0.47). The top signal, rs10462706, was associated with decreased OC risk (OR=0.74, $P=5.54x10^{-10}$) and is in low LD (r^2=0.15, **Supplementary Fig. 8**) with the second strongest signal rs467095. These two variants are 7kb apart and map to intron 13 of *CLPTM1L* and in stratified analysis showed stronger effects in never smokers (P_{het}=0.07 and P_{het}=0.0028, respectively) and never drinkers (P_{het}=0.01 and P_{het}=0.0025, respectively) (**Fig. 3b, Supplementary Fig. 9**). Conditional analyses showed that these SNPs are not completely independent (**Supplementary Table 19**). *TERT* and *CLPTM1L* encode the telomerase reverse transcriptase (*TERT*) and the cleft-lip and palate-associated transmembrane 1-like protein (*CLPTM1L*), respectively. Notably, rs467095 is an esophageal *TERT* eQTL[14] (**Supplementary Table 12**) and is in high LD with rs401681 (OR=1.18, $P=2.1x10^{-7}$, r^2=0.94) a widely studied SNP associated

with risk of several cancers including: lung[15,16], bladder, prostate, cervical, melanoma[17], basal cell[18], esophageal[19], pancreatic[20] and nasopharyngeal[21]. Multiple 5p15.33 variants have been reported to independently influence cancer risk in both an increasing and decreasing fashion. Interestingly, rs401681[A] was associated with an increased OC risk similar to previous melanoma associations, and in an opposite direction to previous lung cancer results[17].

Several variants within the *CDKN2A–CDKN2B* locus (9p21.3) were found to be associated with OC. The lead SNP, rs8181047, is an intronic variant within the CDKN2B1 antisense RNA 1 (*CDKN2B-AS1)* (**Fig. 3c**). rs8181047 is in LD (r^2_{range}=0.6-0.8) with 4 other 9p21.3 variants strongly associated with OC (**Supplementary Fig. 10a**) that in conditional analyses did not show independent associations (**Supplementary Table 20**). The *CDKN2A–CDKN2B* locus contains genes involved in cell-cycle regulation and senescence and has been associated with multiple malignancies including melanoma[22], glioma[23], basal cell[18], breast[24], lung[25], nasopharyngeal[26] and esophageal cancer[27]. Notably, *CDKN2A* is frequently mutated in HPV-negative head and neck cancers[28].

The OC associated variants at 9q34.12 mapped to an intron of *LAMC3,* a laminin involved in cortical development[29]*.* rs928674, the peak signal, showed consistent effects across strata and a weaker association with OPC (*P*=0.003) (**Fig. 3d**). rs928674 is in high LD with 3 other robustly associated 9q34.12 SNPs (r^2_{range}=0.82-0.96, **Supplemental Fig. 10b, Supplementary Table 21**) and is an esophageal mucosa cis-eQTL for a downstream gene *AIF1L* (Allograft Inflammatory Factor 1-Like)[14].

The most prominent finding in the overall and OPC meta-analyses was a large association signal at 6p21.32 within the HLA class II region. The lead variant in both analyses, rs3828805, maps 1.7kb 5' of *HLA-DQB1* (**Fig. 2d** and **Supplementary Fig. 11)** and similar to other 6p21.32 variants (**Supplementary Table 4 and 6**), showed heterogeneity by geographical region (P_{het}=0.007) with no effect in South America (P=0.62). Association analyses of 6p21.32 variants (P<5x10^{-8}) conditioned on rs3828805 did not reveal multiple independent signals (**Supplementary Table 22**), suggesting a common haplotype. To further investigate HLA associations, we imputed classical alleles in 11,436 individuals (>70% Caucasian ancestry) (**Online Methods**). Three classical HLA alleles *DRB1*1301*, *DQA1*0103* and *DQB1*0603* reached P<5x10^{-8} in the overall analysis and were also strongly associated with OPC (**Supplementary Table 23**). These alleles are in high LD (r^2>0.9) and are part of the HLA class II haplotype, *DRB1*1301-DQA1*0103-DQB1*0603,* which is common in Europeans and previously reported to be associated with decreased cervical cancer risk[30]. *DRB1*1301-DQA1*0103-DQB1*0603* was strongly associated with reduced OPC risk (OR=0.59, P=2.7x10^{-9}) and more weakly with OC risk (OR=0.75, P=1.7x10^{-4}). Further conditional analysis on this haplotype and 6p21.32 variants did not reveal evidence of additional independent effects (**Supplementary Table 24**). Given the importance of HPV infection in the etiology of cervical and oropharyngeal cancer[31], we conducted post-hoc analyses to examine the effect of *DRB1*1301-DQA1*0103-DQB1*0603* in a subset of 576 cases with available HPV-status and 3662 controls. *DRB1*1301-DQA1*0103-DQB1*0603* was associated with a strong reduced risk of HPV-positive OPC (OR=0.23, P=1x10^{-6}, n=336) with no significant association in 240 OPC HPV-negative cases (OR=0.75, P=0.16)

(**Table 3**). These results indicate that the class II HLA region is implicated in at least two HPV-driven cancers, namely HPV-positive OPC and cervical cancer. The lack of an association between 6p21.3 SNPs and OPC risk in South America could relate to previous findings that less than 10% of OPC are HPV-positive in this region[4,5]. Moreover, a weaker association with OC could be due to a smaller proportion of these cases being HPV-positive, as well as possibly some misclassified OPC cases, especially for base of the tongue tumors. Further evaluation of the extent and specificity of this HLA effect in HPV-associated cancers is important given the strength of the observed association. This may help elucidate why some individuals are at higher risk of HPV-positive OPC after HPV infection and may also have implications for cancer immunotherapies targeting the HLA class II antigen presentation pathway[32].

In summary, we identified 7 oral and pharyngeal cancer susceptibility loci including a strong HLA signal narrowed to a class II haplotype. Future replication of these findings in an independent population is warranted as well as fine-mapping and functional studies necessary to establish the biological framework underneath these associations.

URLs. R, http://www.r-project.org/; PLINK, https://www.cog-genomics.org/plink2; University of Michigan Imputation Server, https://imputationserver.sph.umich.edu/; Haplotype Reference Consortium, http://www.haplotype-reference-consortium.org/; SNP2HLA, https://www.broadinstitute.org/mpg/snp2hla/; HaploReg v4.1, http://www.broadinstitute.org/mammals/haploreg/haploreg.php; GTEx Portal, http://www.gtexportal.org/home/; LocusZoom, http://locuszoom.sph.umich.edu/locuszoom/; metafor R package http://www.metafor-project.org/doku.php; LDlink, http://analysistools.nci.nih.gov/LDlink/; INHANCE Consortium, http://www.inhance.utah.edu/; OncoArray Network, http://epi.grants.cancer.gov/oncoarray/; GAME-ON, http://epi.grants.cancer.gov/gameon/; GENCAPO, http://www.gencapo.famerp.br

Acknowledgments

We thank Dr. Leticia Fernandez (Instituto Nacional de Oncologia y Radiobiologia, La Habana, Cuba) for her contribution to the IARC ORC multicenter study. We are also grateful to Sergio Koifman (Escola Nacional de Saúde Pública, Rio de Janeiro, Brazil) for his contribution to the IARC Latin America multicenter study (Sergio Koifman passed away in May 2014). We acknowledge all of the participants involved in this research and the funders and support. Genotyping performed at the Center for Inherited Disease Research (CIDR) was funded through the U.S. National Institute of Dental and Craniofacial Research (NIDCR) grant 1X01HG007780-0. Genotyping for shared controls with the Lung OncoArray initiative was funded through the grant X01HG007492-0. Corina Lesseur undertook this work during the tenure of a Postdoctoral Fellowship awarded by the International Agency for Research on Cancer. The funders did not participate in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The University of Pittsburgh head and neck cancer case-control study is supported by National Institutes of Health grants P50 CA097190 and P30 CA047904. The Carolina Head and Neck Cancer Study (CHANCE) was supported by the National Cancer Institute (R01-CA90731). The Head and Neck Genome Project (GENCAPO) was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Grant numbers 04/12054-9 and 10/51168-0). The authors thank all the members of the GENCAPO team. The HN5000 study was funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10034), the views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The Toronto study was funded by the Canadian Cancer Society Research Institute (020214) and the National Cancer Institute (U19 CA148127) and the Cancer Care Ontario Research Chair. The alcohol-related cancers and genetic susceptibility study in Europe (ARCAGE) was funded by the European Commission's 5th Framework Program (QLK1-2001-00182), the Italian Association for Cancer Research, Compagnia di San Paolo/FIRMS, Region Piemonte, and Padova University (CPDA057222).The Rome Study was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) IG 2011 10491 and IG2013 14220 to SB, and Fondazione Veronesi to SB. The IARC Latin American study was funded by the European Commission INCO-DC programme (IC18-CT97-0222), with additional funding from Fondo para la Investigacion Cientifica y Tecnologica (Argentina) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (01/01768-2). The IARC Central Europe study was supported by European Commission's INCO-COPERNICUS Program (IC15-CT98-0332), NIH/National Cancer Institute grant CA92039, and the World Cancer Research Foundation grant WCRF 99A28.The IARC Oral Cancer Multicenter study was funded by: grant S06 96 202489 05F02 from Europe against Cancer; Grants FIS 97/0024, FIS 97/0662, and BAE 01/5013 from Fondo de Investigaciones Sanitarias, Spain; UICC Yamagiwa-Yoshida Memorial International Cancer Study; National Cancer Institute of Canada; Italian Association for Research on Cancer; and the Pan American Health Organization. The coordination of EPIC study is financially supported by the European Commission (DG SANCO) and the International Agency for Research on Cancer.

Competing financial interests: The authors declare no competing financial interests.

Author contributions

P.Brennan and J.D.M. conceived and designed the project. C.L. undertook data harmonization, genotypes quality control, GWAS analysis, imputation and meta-analyses. X.X. performed genotype calling. V.G. and A.C. organized and supervised sample selection and DNA shipments at IARC. A.C. performed replication TaqMan genotying. C.L. and V.G. analyzed data from replication genotyping. C.L. and P.Brennan drafted the first version of the manuscript. B.D., A.F.O, V.W.-F., A.R.N, G.L., M.L., J.E.-N., S.F., P.L., G.J.M, L.R., S.B., J.P., K.K., D.Z., M.J., A.M.M., M.P.C., M.R., W.A., C.C., A.Z., X.C., D.I.C, I.H., D.M., M.V., C.M.H., N.S.-D., E.F., J.L., J.R.G, M.C.W., E.H.T, F.D.N, M.B.C., S.T., R.J.H., W.H.M.P., R.H., G.C., A.S., A.A., O.S., H.B.B.Dm, P.Boffetta and D.A., contributed with reagent/samples/material and reviewed/approved the final manuscript. J.D.M. and C.I.A designed and coordinated the Lung Cancer OncoArray. P.Brennan obtained funding for the project, provided overall supervision and management.

REFERENCES

1. Gillison, M.L., Chaturvedi, A.K., Anderson, W.F. & Fakhry, C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol* **33**, 3235-42 (2015).
2. Chaturvedi, A.K. *et al.* Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of Clinical Oncology* **29**, 4294-4301 (2011).
3. Kreimer, A.R., Clifford, G.M., Boyle, P. & Franceschi, S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev* **14**, 467-75 (2005).
4. Ribeiro, K.B. *et al.* Low human papillomavirus prevalence in head and neck cancer: results from two large case-control studies in high-incidence regions. *Int J Epidemiol* **40**, 489-502 (2011).
5. Lopez, R.V. *et al.* Human papillomavirus (HPV) 16 and the prognosis of head and neck cancer in a geographical region with a low prevalence of HPV infection. *Cancer Causes Control* **25**, 461-71 (2014).
6. Hashibe, M. *et al.* Multiple ADH genes are associated with upper aerodigestive cancers. *Nature genetics* **40**, 707-709 (2008).
7. McKay, J.D. *et al.* A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS genetics* **7**, e1001333 (2011).
8. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv*, 035170 (2015).
9. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* **17**, R122-8 (2008).
10. Hashibe, M. *et al.* Evidence for an important role of alcohol- and aldehyde-metabolizing genes in cancers of the upper aerodigestive tract. *Cancer Epidemiol Biomarkers Prev* **15**, 696-703 (2006).
11. Chang, J.S., Straif, K. & Guha, N. The role of alcohol dehydrogenase genes in head and neck cancers: a systematic review and meta-analysis of ADH1B and ADH1C. *Mutagenesis* **27**, 275-86 (2012).

12. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* **17**, 1418-28 (2014).
13. Carre, C. & Shiekhattar, R. Human GTPases associate with RNA polymerase II to mediate its nuclear import. *Mol Cell Biol* **31**, 3953-62 (2011).
14. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
15. McKay, J.D. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat Genet* **40**, 1404-6 (2008).
16. Wang, Y. *et al.* Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* **40**, 1407-9 (2008).
17. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* **41**, 221-7 (2009).
18. Stacey, S.N. *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet* **41**, 909-14 (2009).
19. Yin, J. *et al.* TERT-CLPTM1L Rs401681 C>T polymorphism was associated with a decreased risk of esophageal cancer in a Chinese population. *PLoS One* **9**, e100667 (2014).
20. Petersen, G.M. *et al.* A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet* **42**, 224-8 (2010).
21. Bei, J.X. *et al.* A GWAS Meta-analysis and Replication Study Identifies a Novel Locus within CLPTM1L/TERT Associated with Nasopharyngeal Carcinoma in Individuals of Chinese Ancestry. *Cancer Epidemiol Biomarkers Prev* **25**, 188-92 (2016).
22. Law, M.H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet* **47**, 987-95 (2015).
23. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* **41**, 899-904 (2009).
24. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-7 (2010).

25. Timofeeva, M.N. *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* **21**, 4980-95 (2012).
26. Bei, J.X. *et al.* A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet* **42**, 599-603 (2010).
27. Wu, C. *et al.* Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. *Nat Genet* **46**, 1001-6 (2014).
28. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576-82 (2015).
29. Barak, T. *et al.* Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat Genet* **43**, 590-4 (2011).
30. Chen, D. *et al.* Genome-wide association study of susceptibility loci for cervical cancer. *J Natl Cancer Inst* **105**, 624-33 (2013).
31. Bouvard, V. *et al.* A review of human carcinogens—Part B: biological agents. *The lancet oncology* **10**, 321-322 (2009).
32. Thibodeau, J., Bourgeois-Daigneault, M.C. & Lapointe, R. Targeting the MHC Class II antigen presentation pathway in cancer immunotherapy. *Oncoimmunology* **1**, 908-916 (2012).

Table 1

Epidemiological and clinical characteristics of cases and controls.

	Cases		Controls	
	N	**%**	**N**	**%**
Total	6034		6585	
Tumor site				
Oral cavity	2990	49.55		
Oropharynx	2641	43.77		
Hypopharynx	305	5.05		
Overlapping	73	1.21		
Other	25	0.41		
Geographic Region				
Europe	2499	41.42	2928	44.46
North America	2549	42.24	2522	38.30
South America	986	16.34	1135	17.24
Sex				
Male	4527	75.02	4325	65.68
Female	1507	24.98	2260	34.32
Age, years				
=<50	1315	21.79	1355	20.58
50-<60	2006	33.24	1954	29.67
60-<70	1748	28.97	1983	30.11
>=70	964	15.98	1293	19.64
Unknown	1	0.02	0	0.00
Smoking Status				
Never	1057	17.52	2508	38.09
Former	1792	29.70	2263	34.37
Current	2623	43.47	1466	22.26
Unknown	562	9.31	348	5.28
Drinking Status				
Never	820	13.59	1199	18.21
Ever	4840	80.21	4840	73.50
Unknown	374	6.20	546	8.29

Table 2

Lead genome-wide significant SNP per validated locus from the regional meta-analyses of OC[a] and pharynx cancer combined, as well as OC[a] and OPC[a] separately.

Region	SNP	chr:pos[b]	Gene	EA/ OA[c]	Info (Rsq)[d]	AF[e] case/ control	OR	*P*	P_{het}
Oral and pharyngeal cancer									
4q23	rs1229984	4:100239319	*ADH1B*	A/G	Geno	0.03/0.06	0.56	$2.29x10^{-15}$	0.002
6p21.32	rs3828805	6:32636120	*HLA-DQB1*	C/T	0.88	0.75/0.72	1.28	$3.35x10^{-13}$	0.007
10q26.13	rs201982221	10:126157446	*LHPP*	D/I	Geno	0.03/0.02	1.67	$1.58x10^{-9}$	0.50
11p15.4	rs1453414	11:5829084	*OR52N2/TRIM5*	C/A	Geno	0.23/0.20	1.19	$4.78x10^{-8}$	0.55
Oral cancer									
2p23.3	rs6547741	2:27855924	*GPN1*	A/G	0.98	0.50/0.54	0.83	$3.97x10^{-8}$	0.34
4q23	rs1229984	4:100239319	*ADH1B*	A/G	Geno	0.03/0.06	0.57	$1.09x10^{-9}$	0.02
5p15.33	rs10462706	5:1343794	*CLPTM1L*	T/C	0.97	0.12/0.15	0.74	$5.54x10^{-10}$	0.84
9p21.3	rs8181047	9:22064465	*CDKN2B-AS1*	A/G	Geno	0.29/0.24	1.24	$3.80x10^{-9}$	0.37
9q34.12	rs928674	9:133952024	*LAMC3*	G/A	0.89	0.14/0.12	1.33	$2.09x10^{-8}$	0.88
Oropharyngeal cancer									
4q23	rs1229984	4:100239319	*ADH1B*	A/G	Geno	0.02/0.06	0.55	$8.53x10^{-9}$	0.05
6p21.32	rs3828805	6:32636120	*HLA-DQB1*	C/T	0.88	0.75/0.7	1.37	$2.21x10^{-12}$	0.07

[a] OC=oral cancer, OPC=oropharyngeal cancer; [b] SNP position according to NCBI genome build 37 (Hg19); [c] EA=Effect allele; OA=other allele; [d] Geno=genotyped, SNP, INFO, R^2 is the average across imputation batches; [e] AF= allele frequency of the effect allele

Figure 1

Genome-wide associations meta-analyses results. The red line represents P=5x10^{-8}. The y-axis represents the –log10 P-values. (a) Overall OC and pharyngeal cancer 6,034 cases and 6585 controls (b) Oral cancer analysis with 2,990 cases and 6,585 controls (c) Oropharyngeal cancer analysis with 2,641 cases and 6,585 controls. Loci with GWA significant SNPs and technically validated are tagged with genomic location.

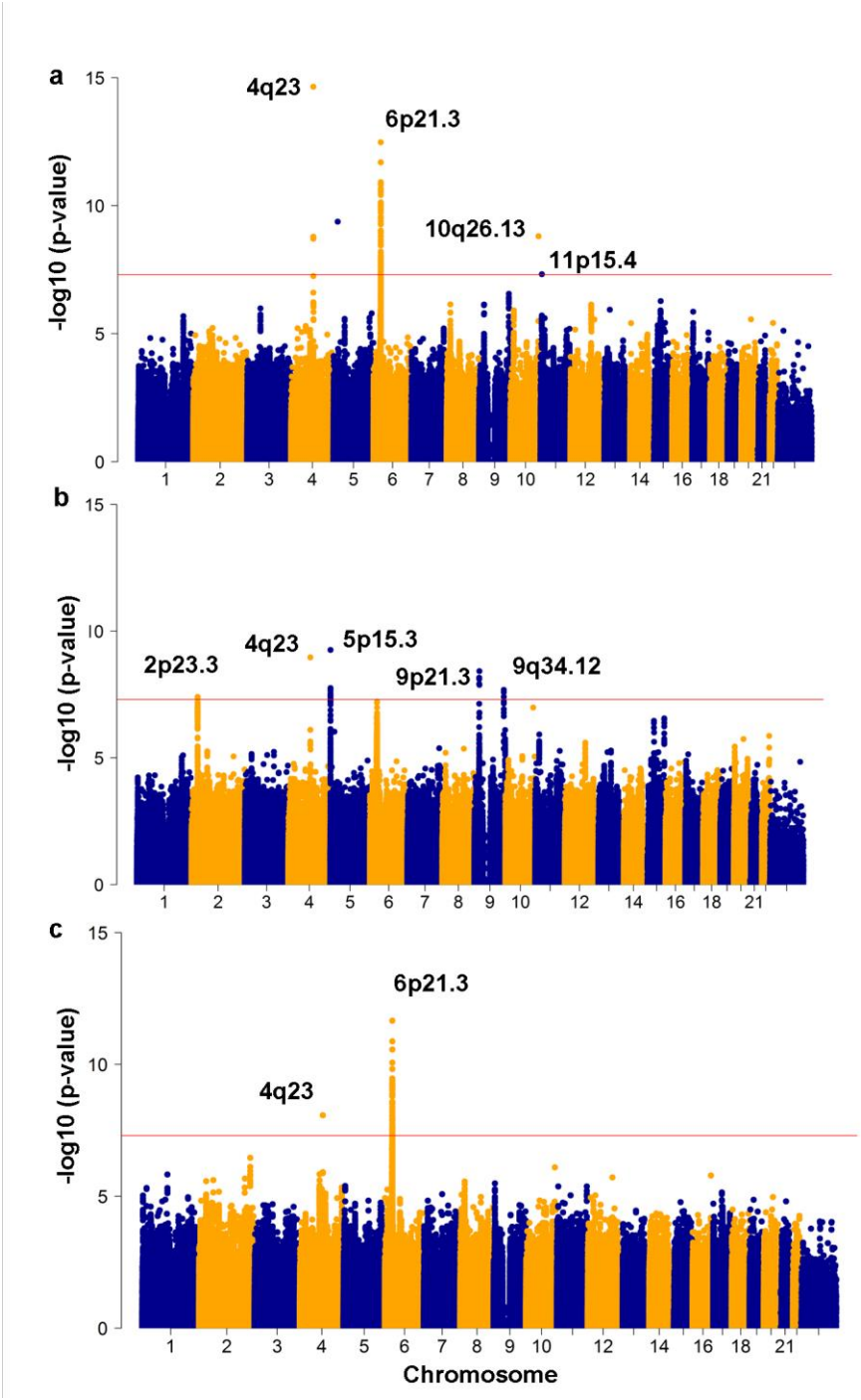

Figure 2

Forest plots of odds ratios for the lead SNP at each genome-wide significant loci in the overall oral and pharyngeal cancer meta-analysis. (a)4q23, rs1229984 (b)10q26.13, rs20198222 (c)11p15.4 rs1453414. (d)6p21.32 rs3828805. EAF=effect allele frequency in 6585 controls. Effect allele in square brakets. OC=oral cancer; OPC=oropharynx cancer. Plot data in Supplementary Table 25.

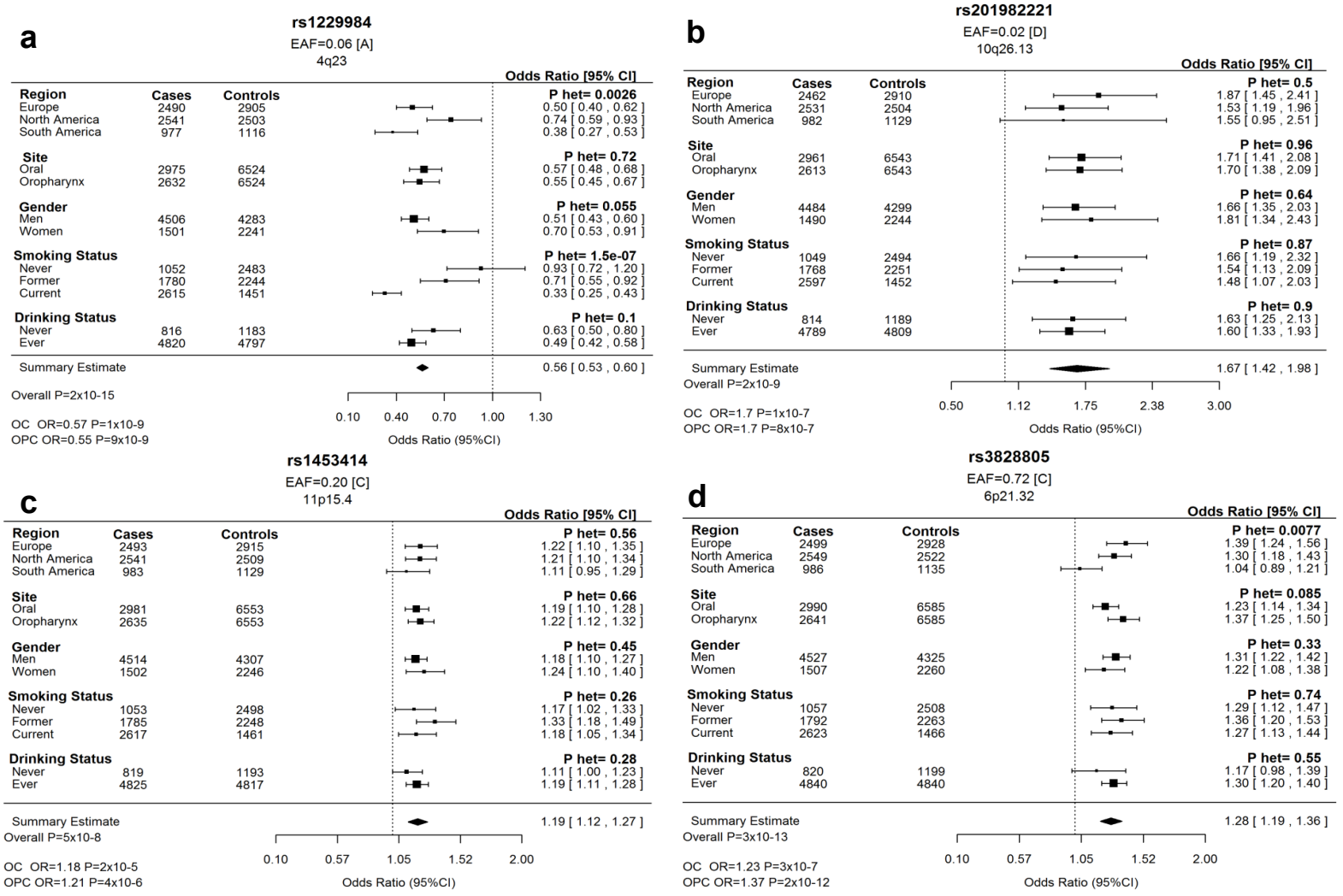

Figure 3

Forest plots of odds ratios for the lead SNP at each genome-wide significant loci in the oral cancer (OC) meta-analysis. (a)2p23.3 rs6547741 (b)5p15.33, rs10462706 (c)9p21.3, rs8181047 (d)9q34.12, rs928674. EAF=effect allele frequency in 6585 controls. Effect allele in square brakets. Overall=oral and pharyngeal cancer; OPC=oropharynx cancer. Plot data in Supplementary Table 26.

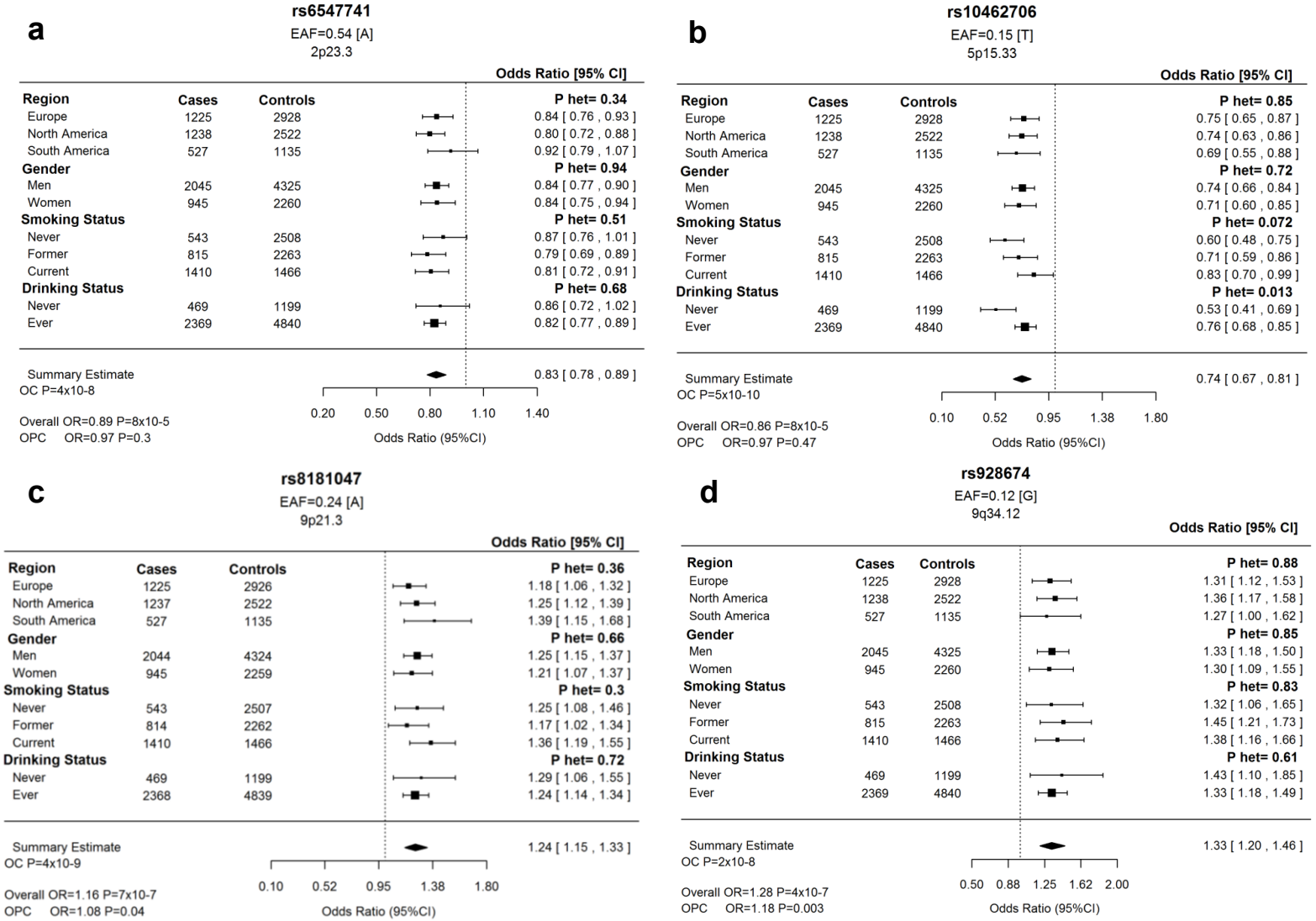

Table 3

Associations of *DRB1*1301-DQA1*0103-DQB1*0603* haplotype in individuals with >70% Caucasian ancestry

							Meta-analysis[c]	
	haplotype[a] case/control	**case/ control**	**HF[b] case**	**HF[b] control**	**OR**	**P**	**OR**	**P**
Oral and pharynx cancer							0.68	3.32×10^{-10}
Europe	207/422	2497/2928	0.04	0.07	0.60	4.04×10^{-8}		
North America	207/276	2342/2329	0.04	0.06	0.74	1.68×10^{-3}		
South America	74/101	647/727	0.06	0.07	0.86	0.35		
Oral cancer							0.75	1.72×10^{-4}
Europe	106/422	1231/2928	0.04	0.07	0.60	1.52×10^{-5}		
North America	128/276	1135/2329	0.06	0.06	0.92	4.67×10^{-1}		
South America	41/101	351/727	0.06	0.07	0.80	0.26		
Oropharynx cancer							0.59	2.73×10^{-9}
Europe	84/422	1098/2928	0.04	0.07	0.57	2.69×10^{-5}		
North America	72/276	1119/2329	0.03	0.06	0.52	3.49×10^{-6}		
South America	31/101	216/727	0.07	0.07	1.05	0.81		
Oropharynx cancer by HPV status[d]								
HPV-positive	11/505	336/3686	0.01	0.07	0.23	1.6×10^{-6}		
HPV-negative	25/505	240/3686	0.05	0.07	0.75	0.16		

[a] Number of copies of the haplotype in cases and controls

[b] Haplotype frequency calculated as total number of copies of haplotype in the population (haplotype copies/2n).

[c] Fixed-effects meta-analysis of regional associations adjusted for age, sex and eigenvectors.

[d]HPV-status available in a subset of cases from ARCAGE, EPIC, CHANCE and Pittsburgh studies